



[AI FACTORY 공모전]

2023 연구개발특구 AI SPARK 챌린지

- 공기압축기 이상 판단 -

대회 기간 : 23.04.03 ~ 23.04.21



김영준

이국현

김수현

오세진



TA
VE



01 대회 소개

- 01 대회 소개
- 02 시계열 데이터

02 EDA

- 01 데이터 input & output
- 02 데이터 시각화
- 03 데이터 전처리

03 모델 구축

- 01 LSTM-AE
- 02 Transformer
- 03 Anomaly Transformer

04 결론

- 01 고찰 및 결론

2023 연구개발특구 AI SPARK 챌린지 :

공기압축기 이상 판단

디지털트윈 및 스마트팩토리 시스템 기반의 산업현장 혁신 솔루션 도출

제4회 연구개발특구 인공지능 경진대회

AI SPARK CHALLENGE

연구개발특구진흥재단은
디지털트윈 및 스마트팩토리 시스템 기반의
산업현장 혁신 솔루션을 도출하고자 합니다.

참가대상 >
데이터 활용 및 인공지능을 통한 문제해결에
관심있는 누구나 ※개인 또는 팀(최대4명)으로 참여 가능

접수기간 >
3월 17일(금)부터 종료일까지 상시 접수 가능

대회기간 >
2023. 4.03(월) 10:00~4.21(금) 18:00

총 상 금 > 800만원

구분	상금
1등(1팀)	500만원
2등(1팀)	200만원
3등(1팀)	100만원

대회안내 >

QR코드 링크정보
bit.ly/aisparkchallenge4th

문의처 >
인공지능팩토리
cs@aifactory.page / 042-710-6451
연구개발특구진흥재단 기술창업지원팀
hjkim@innopolis.or.kr / 042-865-8983

주최
과학기술정보통신부 INNOPOLIS 연구개발특구진흥재단

주관/운영
파워팩토리연구소 AiFrenz AIFRICA AIFactory

대회 목표

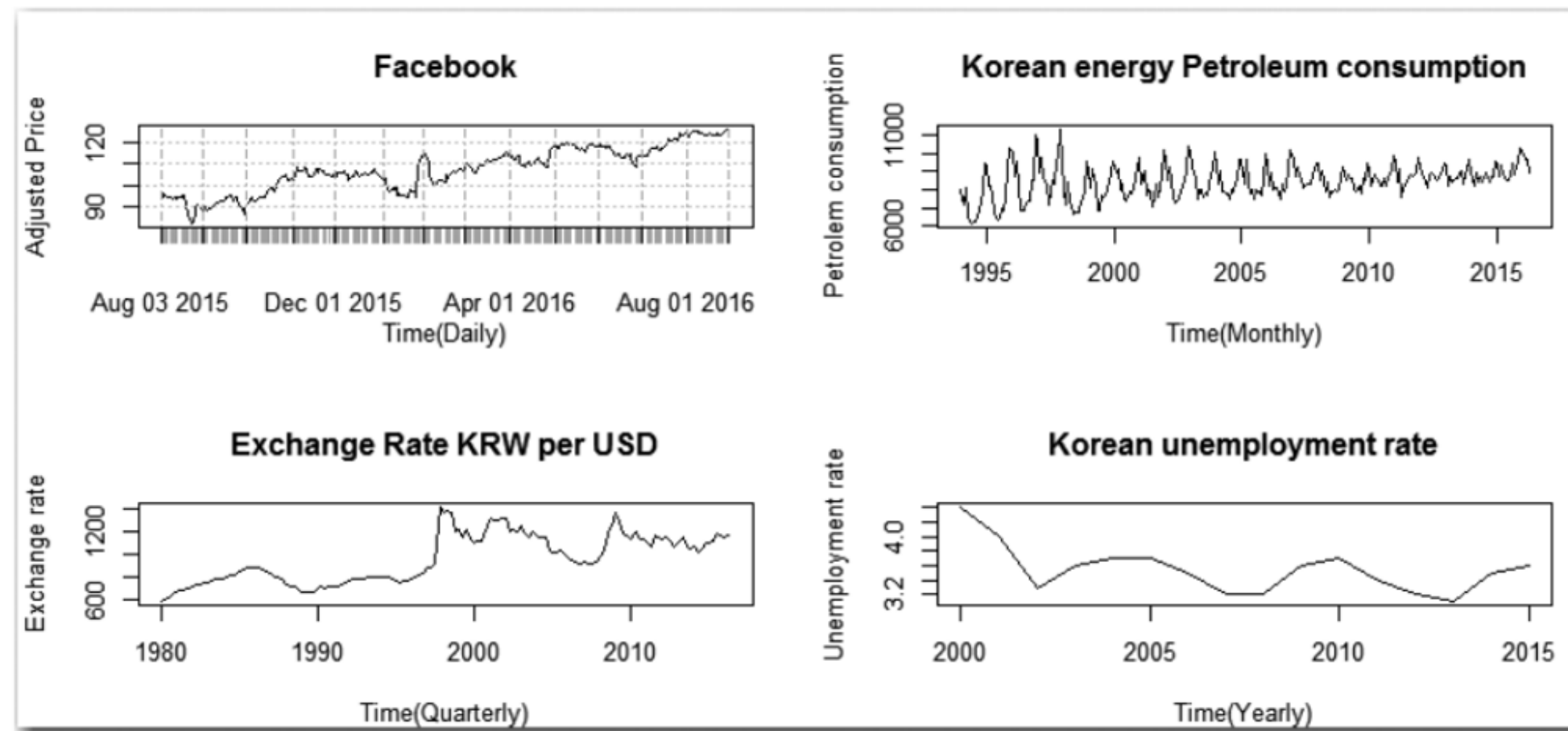
산업기기의 이상 전조증상을 예측하여 기기
고장 예방과 사고 예방을 하는 모델 개발

POINT

- ✓ 비지도 학습(Unsupervised learning) 방식
- ✓ 향후 실시간 판정에 활용 가능해야하고, 입력된 데이터를 0(정상) 또는 1(이상)로 이진 분류하는 이상 판단 모델
- ✓ 입력 시계열 데이터를 판정할 수 있는 모델

시계열 데이터

시계열 데이터는 시간 순서대로 나열된 데이터를 의미합니다.
즉, 일정한 시간 간격으로 수집된 데이터로서, 시간에 따른 변화를 관찰할 수 있습니다.



시계열 데이터 분석 목적

시계열이 갖고 있는 법칙성을 발견하여 이를 모형화하고, 추정된 모형을 통하여 미래의 값을 예측 하는 것

INPUT, OUTPUT

Train : 약 2400개
Test : 약 7400개

INPUT

air_inflow	air_end_temp	out_pressure	motor_current	motor_rpm	motor_temp	motor_vibe	type
1.59	41	0.7	20.53	1680	58.67	2.93	0
2.97	59.28	0.7	38.4	3142	74.91	3.75	0
1.91	45.29	0.7	24.73	2023	62.48	3.12	0
2.37	51.33	0.7	30.63	2506	67.84	3.39	0
1.9	45.21	0.7	24.65	2017	62.41	3.12	0
2.22	49.34	0.7	28.69	2347	66.08	3.3	0
2.92	58.7	0.7	37.84	3096	74.4	3.72	0
3.23	62.7	0.7	41.75	3416	77.96	3.9	0
2.11	47.93	0.7	27.3	2234	64.82	3.24	0
2.88	58.08	0.7	37.23	3046	73.84	3.69	0
2.46	52.58	0.7	31.85	2606	68.96	3.45	0
1.34	37.78	0.7	17.38	1422	55.8	2.79	0
2.35	51.09	0.7	30.4	2487	67.63	3.38	0

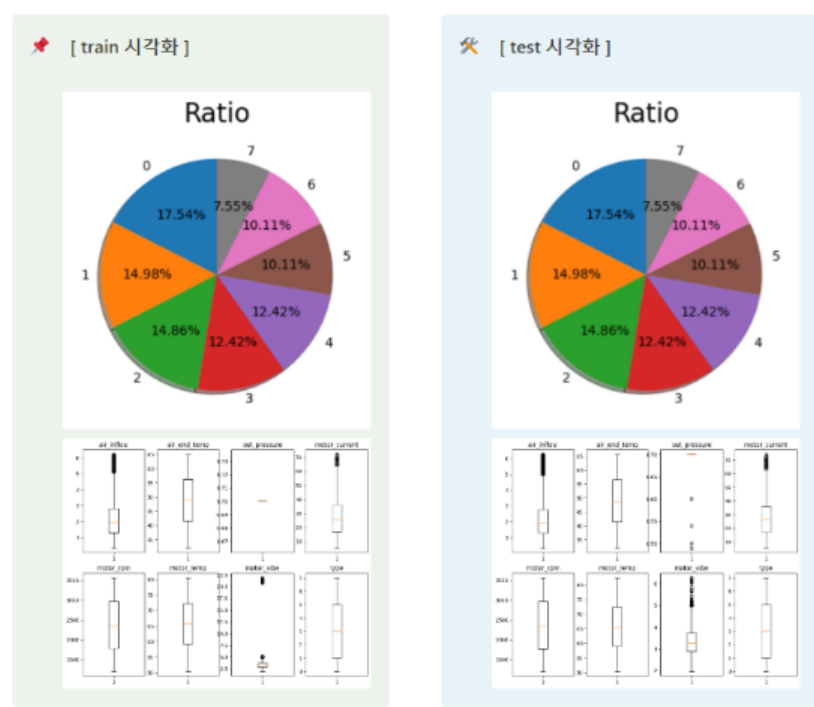
OUTPUT

type	label
0	0
0	0
0	0
0	0
0	0
0	0
0	0
0	0
0	0
0	0
0	0
0	0
0	0
0	0
0	0

INPUT 공기 흡입 유량, 공기 말단 온도, 토출 압력, 모터 전류, 모터 회전수 등 총 8개의 feature를 갖는 시계열 데이터**OUTPUT** type별 label 예측 결과에 따라 0(정상) 또는 1(이상)을 출력

01

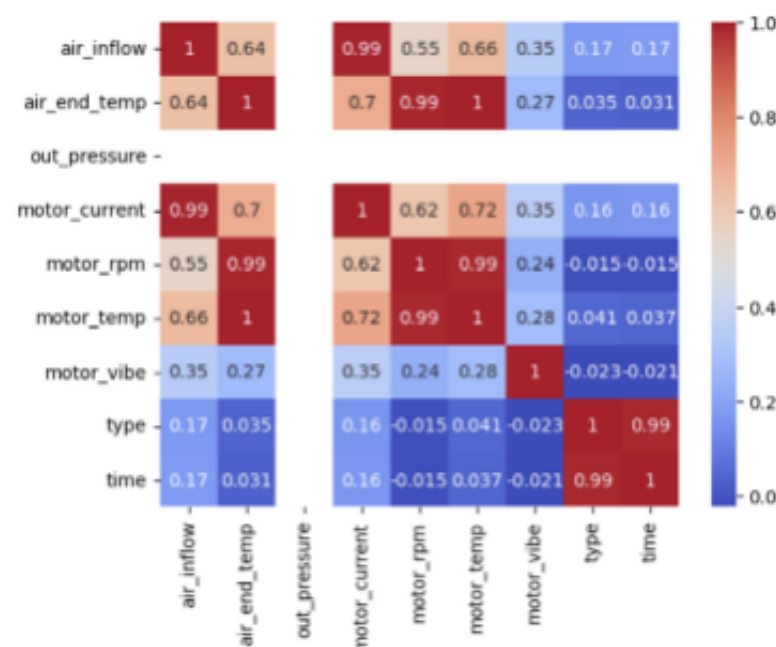
데이터 분포 확인



- ✓ type별 분포량 -> 데이터 불균형 확인
- ✓ type별 데이터 분포와 이상치 확인

02

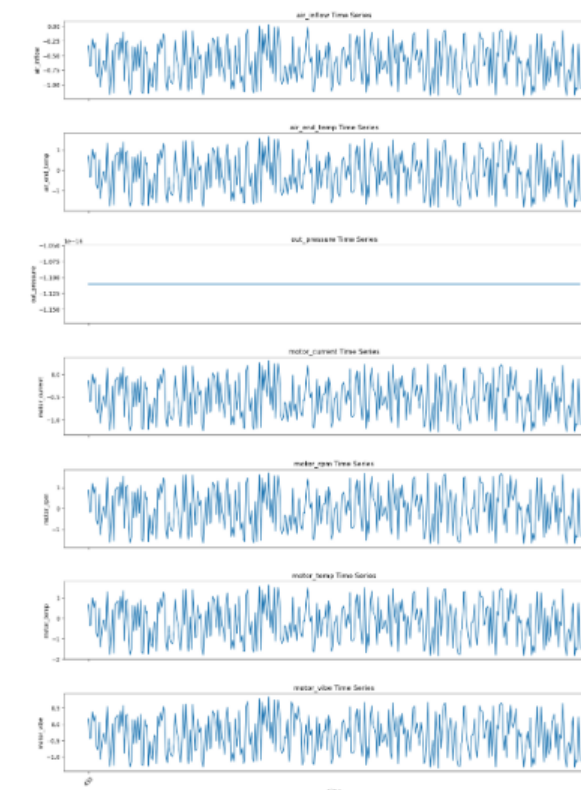
Feature 연관성 파악



- ✓ 독립변수들 간 강한 상관관계 파악
↓
다중공선성 해결

03

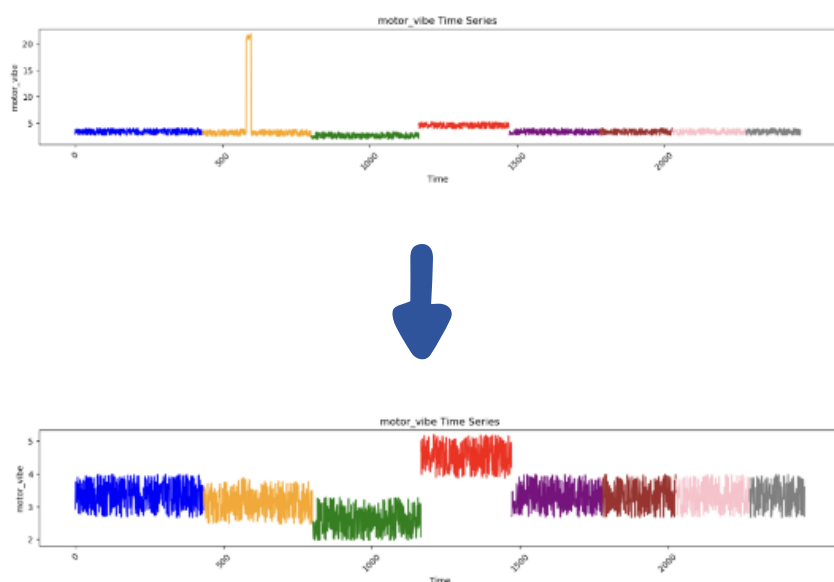
시계열 데이터 패턴 파악



- ✓ x 축 : 시간, y축 : type
- ✓ type별 패턴 비교

01

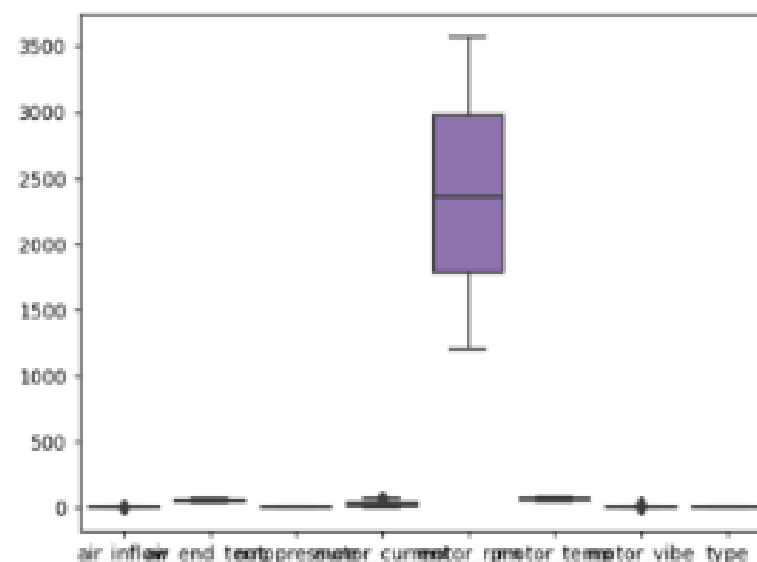
데이터 범위 조절



- ✓ 정상 데이터 내 특이값 확인
- ↓
- 특이값을 type의 최댓값과 최솟값 사이의 값으로 변경

02

Feature Scaling



- ✓ feature 범위 확인 후 scaling
- ↓
- 학습 시 범위에 따른 가중치 편향 최소화

03

Feature 제거/추가

type	hp
0	30
0	30
0	30
0	30
0	30
0	30
0	30
0	30
0	30
0	30

- ✓ type feature 제거 -> 마력으로 대체
- ✓ hp feature 추가 -> type에 따라 마력 특징 맞춤

LSTM-AutoEncoder(AE)


F1-SCORE

Final : 0.712 (Public : 0.715)

LSTM-AE 한계

01

여전히 long-term dependencies 문제

- ✓ LSTM도 RNN의 일종
- ✓ 긴 sequence data 처리 불가

02

Dataset 한계

- ✓ # train size 적음 (약 2400개)
- ✓ test data는 정상 및 비정상 데이터
- ✓ train에 overfitting 방향만 존재

03

더 크고 성능 좋은 모델 필요성

- ✓ parameter가 큰 모델 구축 필요
- ✓ 효과적인 pre-trained model 구축 필요

INSIGHT

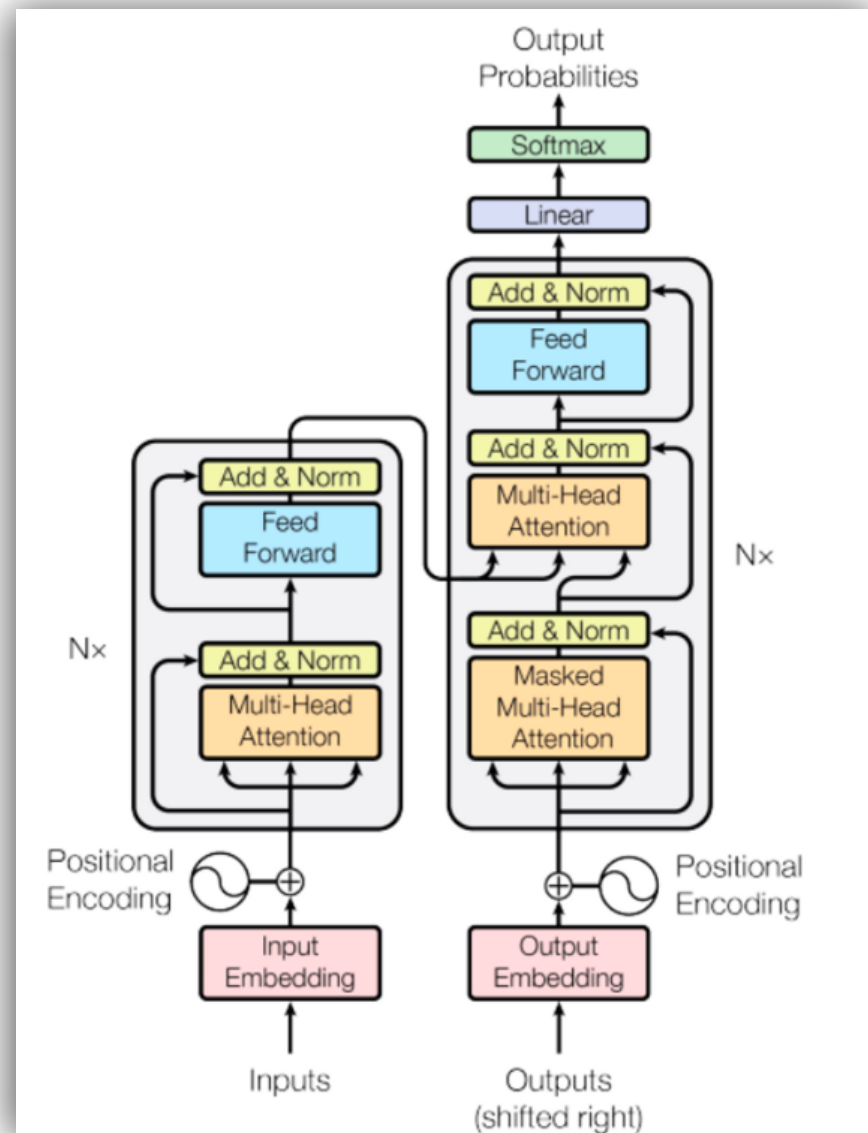
Attention이 있는 Transformer를 사용

01

02

03

04



Transformer Architecture

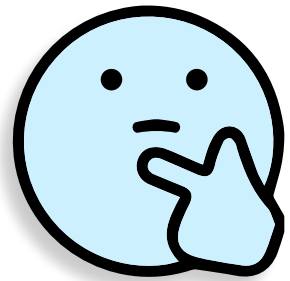
소개

- ✓ Transformer는 2017년에 제안된 딥러닝 아키텍처
- ✓ LSTM에 비해 긴 시퀀스(시계열 batch)를 더 효과적으로 처리

주요 개념

- ✓ **Attention 메커니즘:**
입력 시퀀스의 각 요소 간의 관계를 고려하여 가중치를 학습함으로써, 전체 시퀀스에 대한 정보를 더 잘 파악
- 이를 활용하여 비지도 학습 시계열 데이터 작업에서도 데이터의 복잡한 상관 관계를 파악하여 다양한 시계열 작업에 유용하게 활용 가능

Transformer 계열 모델 선정



TRANSFORMER

TranAD

(2022, VLDB)



두 개의 디코더를 사용



Adversarial 훈련 과정 복잡,
학습이 수렴하기 어려움 존재

**Anomaly
Transformer**

(2022, ICLR)



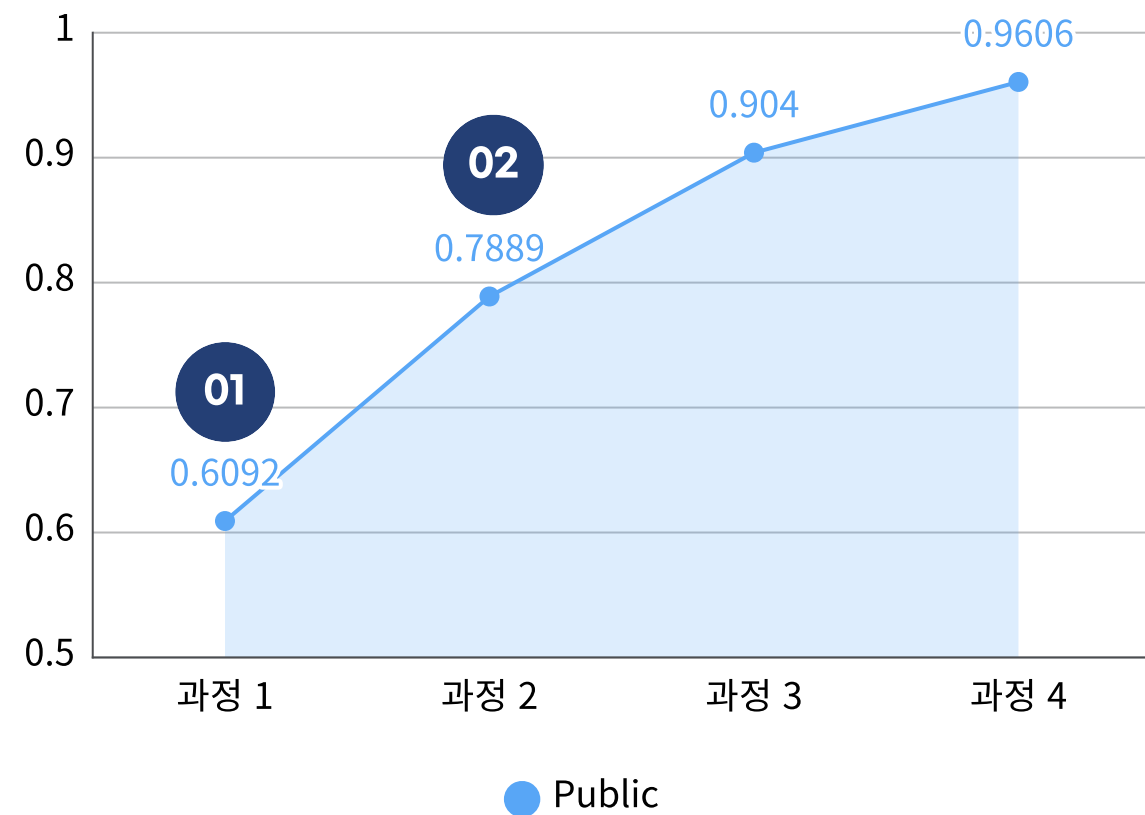
지역적 시계열 특징과 전반적 시계열 특징을
결합하여 더욱 정확한 이상치 탐지를 제공



두 가지 시계열 특징에 대한 이론적 이해 필요

Anomaly Transformer

제시된 downstream task에 맞게 model 개선을 진행하였고, 이 과정은 네 가지 주요 단계로 구성되어 있다.



01

Vanila Anomaly Transformer 적용

- ✓ task에 맞게 model의 input과 output 채널 수정
- ✓ 설비 번호(type)별로 model train & test 진행

INSIGHT

train dataset이 매우 작아 model 성능이 낮게 나온다고 판단하여 Pre-trained model을 활용하기로 결정

02

Pre-trained model 사용

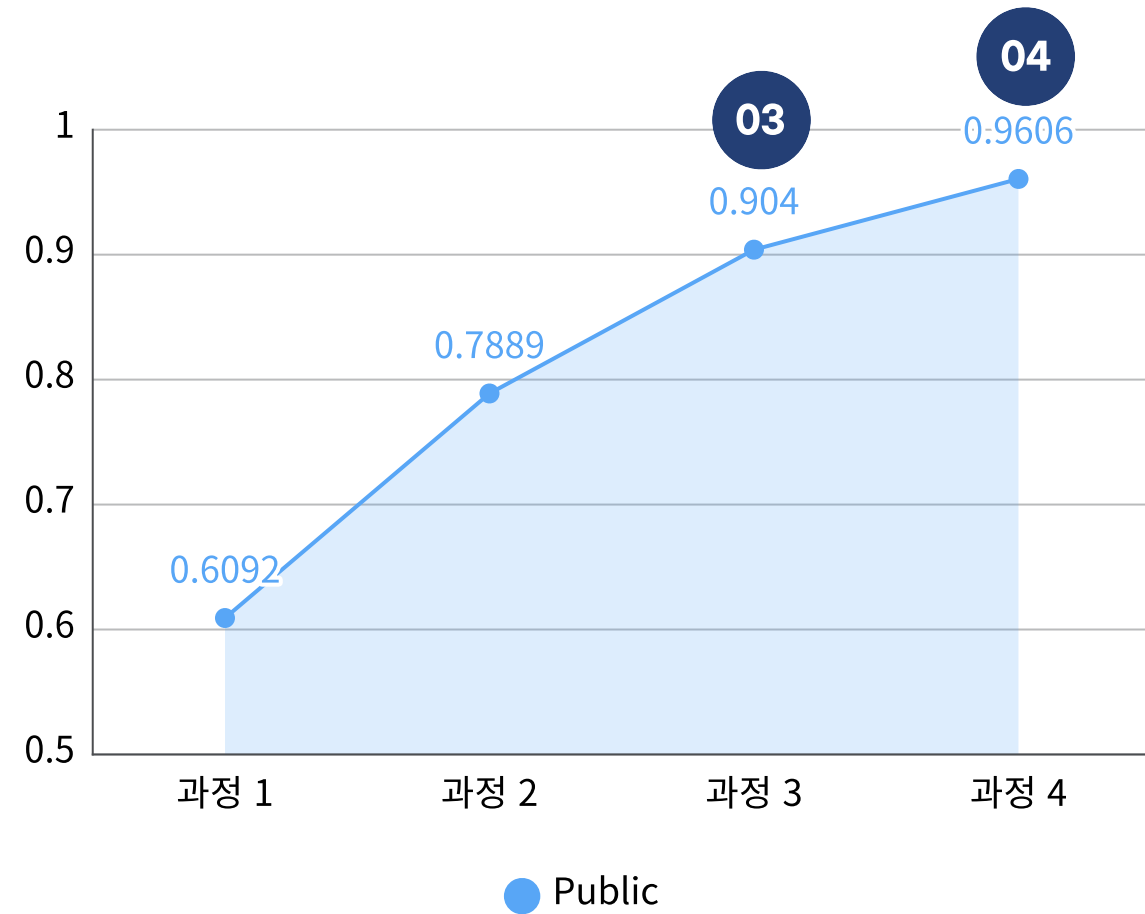
- ✓ 산업용 시스템의 이상감지 문제를 해결하는데 사용되며 task 데이터와 같은 시계열 데이터인 PSM dataset을 pre-train한 model 사용
- ✓ pre-trained model을 위해 PSM dataset의 차원을 PCA를 통해 축소

INSIGHT

1. Pre-train & Fine-tune 방법을 사용하려면 두 method의 dataset의 feature 개수를 통일해야하는 한계가 존재함
2. PSM dataset pre-trained model을 진행했지만 성능 향상X

Anomaly Transformer

제시된 downstream task에 맞게 model 개선을 진행하였고, 이 과정은 네 가지 주요 단계로 구성되어 있다.



03

Feature 추가 및 전체 Dataset 학습

- ✓ 'HP'(마력)이라는 feature를 추가
- ✓ 설비 번호(type)별로 학습시키지 않고 전체 train dataset을 학습시킴

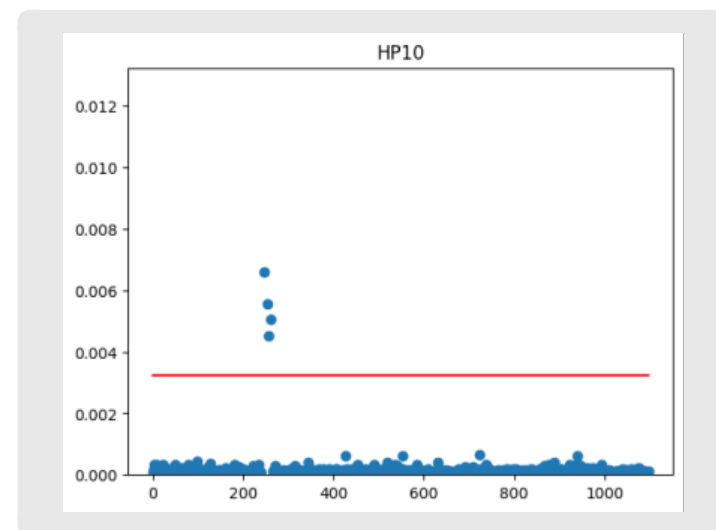
INSIGHT

설비 번호(type)별로 학습 시 데이터 부족으로 성능이 향상되는 데에 한계가 있다고 판단.
 ➔ 데이터의 특성 중 하나인 HP(마력)을 이용하고자 함

04

최종 Model

- ✓ 전체 dataset을 'HP'(마력)별로 나눠서 학습 및 추론
- ✓ 'HP'(마력)별 hyper-parameter tuning하여 성능 개선
- ✓ 시각화를 통해 threshold 조정



고찰 및 결론

인지

도메인 이해

전혀 생소한 분야의 도메인이었지만, 특징들을 분석해보며 차근차근 이해를 쌓아나갈 수 있었다.

학습 과정에서 도메인에 대한 지식을 끊임없이 쌓아가며, 이를 바탕으로 데이터의 특성을 파악하고 의미 있는 feature들을 추출할 수 있었다. 이 과정은 앞으로 새로운 도메인에 도전할 때 큰 도움이 될 것 같다.

경험

태스크 분석

이번 공모전에서 '비지도 학습', '이상치 탐지' 그리고 '시계열 데이터'라는 주요 task에 집중하여 분석을 수행했다. **각각의 task를 깊이 이해하고 적절한 해결 전략을 세우며 과정을 진행했다.** 이를 통해 다양한 도메인에서의 문제해결 능력을 키울 수 있었으며, 해당 task와 관련된 인사이트를 넓힐 수 있는 계기가 되었다.

확장

모델 동향 파악

머신러닝 모델부터 LSTM-AE 그리고 최신 Transformer 계열 모델까지 다양한 딥러닝 모델의 발전 동향을 살펴보았다. 이러한 모델들을 실제 task에 적용하고 구현해보며, 성능을 직접 측정하는 시간을 가졌다. **이 경험을 바탕으로 앞으로 다양한 딥러닝 프로젝트에서 적절한 모델 선택과 성능 최적화에 대한 판단력이 향상될 것으로 기대한다.**



AI Factory 공기압축기 이상 판단 공모전

QnA



Notion 링크 :

<https://skipper0527.notion.site/TAVE-36ca711c05284e41b4dc8b79d0c7f75d>



TA
VE